DOCUMENT RESUME

ED 201 456 RC 012 684

AUTHOR Cervantes, Robert A.; Bernal, Helen Hazuda
TITLE A Comparative Analysis of English and Spanish Reading Performance of Mexican American Students.
INSTITUTION D.C. Development Associates, San Antonio, Tex.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Jun 76
GRANT OEC-0-72-5309
NOTE 36p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Bilingual Education; Bilingual Students; Comparative Analysis; Comparative Testing; Elementary Secondary Education; *English; Grade 3; Grade 5; Grade 8; Mexican American Education; *Mexican Americans; Program Effectiveness; *Reading Achievement; *Reading Tests; *Spanish; Test Reliability; Test Validity
IDENTIFIERS Guidance Testing Assoc Inter American Test Reading; *Texas (South)

ABSTRACT

A South Texas survey conducted in 1975 investigated the reading performance of Mexican American students enrolled in a bilingual program to determine whether or not students achieved significantly different reading scores on parallel Spanish and English versions of an appropriate test (Guidance Testing Associates Inter-American Test of Reading). Designed especially for the control group so that the socio-cultural and dialectical sources of bias in either language version should be negligible, the test was given in April and May to a selection of students in grades three, five, and eight in order to determine whether the fit, or lack thereof, between English and Spanish scores was attenuated across grade levels. Mean scores for all students were found to be significantly higher on the English than the Spanish subtest. These differences were interpreted as casting considerable doubt on the commonly held assumption that Spanish language tests are "ipso facto" superior to English language tests for Mexican American, Spanish-surnamed bilingual pupils. Implications were that a re-examination of bilingual programs with respect to their philosophical premises and pedagogical practices appears warranted, and that additional empirical studies are needed in the area of Spanish-English language dynamics with respect to student achievement outcomes. (JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED2014456

D. C. DEVELOPMENT ASSOCIATES, INC.
MANAGEMZNT AND GOVERNMENTAL CONSULTANTS
MILAM BUILDING · SUITE 1400
SAN ANTONIO, TEXAS 78205
512/224-2319

A Comparative Analysis of English and Spanish Reading Performance of Mexican American Students

by
Robert A. Cervantes, Ph.D.
Principal Investigator
Helen Hazuda Bernal, Ph.D.
Consultant

Submitted to:

National Institute of Education
Experimental Schools Program
1200 19th Street
Washington, D.C. 20208

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

RC 012684

The contents reported herein were supported by grant #OEC-0-72-5309 from the National Institute of Education, Experimental Schools Program, of the Department of Health, Education, and Welfare. The opinions expressed do not necessarily reflect the position or policy of the National Institute of Education and no official endorsement should be inferred. The authors wish to thank Drs. William C. Theimer, Earl Jones, Ernest Bernal, and Barry Thompson for their review and comments on the initial draft of this report.

June 1976

HOME OFFICE — 1521 NEW HAMPSHIRE AVENUE, N. W., WASHINGTON, D. C. 20036 • 202/332-5293
DENVER — 180 COOK STREET, SUITE 207, DENVER, COLORADO 80206 • 303/321-2272
HOUSTON — 3220 LOUISIANA, SUITE 203, HOUSTON, TEXAS 77006 • 713/522-2254
LOS ANGELES — 2140 WEST OLYMPIC BLVD., SUITE 534, LOS ANGELES, CALIFORNIA 90006 • 213/487-5740

Table of Contents

	page
List of Tables	iii
List of Figures	iv
Introduction	1
Research Questions	2
Methodology	3
Research Design and Sample	3
Instrumentation	4
Guidance Testing Associates Reading Tests	4
Comprehensive Tests of Basic Skills	5
Procedure	6
Analytical Techniques	6
Findings Related to Research Questions	6
Grade 3	6
Total Battery Scores	7
Vocabulary Scores	9
Level of Comprehension	11
Speed of Comprehension	13
Summary and Discussion	15
Grade 5	16
Total Battery Scores	17
Vocabulary Scores	19
Level of Comprehension	21
Speed of Comprehension	23
Summary and Discussion	25
Comparison of Findings for Grades 3 and 5	26
Correlations of Reading Tests	26
English-Spanish GTA Correlations	26
English GTA-CTBS Correlations	27
Discussion and Implications	28
Bibliography	30

List of Tables

		page
Table 1:	ANOVA Source Table for Spanish Total Battery by Group and Sex: Grade 3	7
Table 2:	ANOVA Source Table for Spanish Vocabulary and Means by Group and Sex: Grade 3	9
Table 3:	ANOVA Source Table for Spanish Level of Comprehension by Group, Sex: Grade 3	11
Table 4:	ANOVA Source Table for Spanish Speed of Comprehension by Group, Sex: Grade 3	13
Table 5:	ANOVA Source Table for English Total Battery by Group and Sex: Grade 5	17
Table 6:	ANOVA Source Table for Spanish Vocabulary by Group and Sex: Grade 5	19
Table 7:	ANOVA Source Table for Level of Comprehension Spanish by Group, Sex, and Comprehension English: Grade 5	21
Table 8:	ANOVA Source Table for Speed of Comprehension Spanish by Group, Sex, and English Comprehension: Grade 5	23
Table 9:	English-Spanish GTA Correlations	26
Table 10:	English GTA-CTBS Correlations: Grades 3, 5, and 8	27

List of Figures

		page
Figure 1:	Total Battery Means by Group and Sex: Grade 3	8
Figure 2:	Vocabulary Subtest Means by Group and Sex: Grade 3	10
Figure 3:	Level of Comprehension Subtest Means by Group and Sex: Grade 3	12
Figure 4:	Speed of Comprehension Subtest by Group and Sex: Grade 3	14
Figure 5:	Total Battery Means by Group and Sex: Grade 5	18
Figure 6:	Vocabulary Subtest Means by Group and Sex: Grade 5	20
Figure 7:	Level of Comprehension Subtest Means by Group and Sex: Grade 5	22
Figure 8:	Speed of Comprehension Subtest Means by Group and Sex: Grade 5	24

INTRODUCTION

The issue of testing practices with respect to Mexican American students is both critical and controversial for test results and their use vitally affect the life chances of these students. Increasingly, researchers have pointed out testing issues related to cultural biases, limitations, and detrimental use and effect with regard to minority groups in general and Mexican American students in particular. (1, 4, 6, 10, 11, 13, 14, 17) Related criticisms have likewise been directed toward the use of standardized English reading tests with monolingual and bilingual Mexican American students which have underscored the problems of test validity raised by language interference between English and Spanish in terms of semantics, dialectical elements, and phonology or syntax. (7, 8)

Since much of the recent criticism concerning the education of Mexican American students has focused on the need for Spanish as the language of instruction to improve achievement -- with the attendant proliferation of widely varying bilingual-bicultural programs of instruction -- careful attention must also be directed toward the problems of testing the achievement of students enrolled in such programs in both English and Spanish. This is an area that, with few exceptions, has received little investigation and may offer important information with respect both to the validity and appropriateness of using Spanish or English versions of tests administered to Mexican American students and to the effectiveness or actual learning outcomes of particular types of bilingual programs of instruction.

The purpose of this study, therefore, was to investigate one area of achievement, namely reading performance, of Mexican American students enrolled in a bilingual program to determine whether or not these Mexican American students achieved significantly different reading scores on parallel Spanish and English versions of an appropriate test. Two points should be noted at the outset. First, the test selected for this study - the Guidance Testing Associates (GTA) Inter-American Test of Reading - was specifically designed for South Texas Mexican American pupils so that the socio-cultural and dialectical sources of bias in either language version should be negligible. Secondly, little can be said about the specific instructional components of the bilingual program in which these students were enrolled other than that the school district sampled labelled the program bilingual and that the most intensive Spanish language instruction occurred in grades K through 2.

With respect to inferences and conclusions drawn in later portions of this report, the reader should keep in mind that bilingual programs can follow one of several models with respect to Spanish language proficiency. Model I is a program in which initial instruction occurs in Spanish, but once basic language-

conceptual skills are developed and transferred to English, Spanish is dropped and is replaced by English as the language of instruction. Model II is a program in which initial instruction occurs in Spanish, but as English skills are developed, instruction continues in both languages so that skills developed in the Spanish language are reinforced or maintained, while English skills receive intensive emphasis. A third type program, Model III, seeks to establish continued growth in both languages by providing instruction in Spanish and English over an extended time period, preferably through high school. The actual learning outcomes with respect to formal Spanish language proficiency can be expected to vary considerably depending on which of these models served as the basis for instruction. Systematic observation revealed that classrooms reflected the spectrum of little or no instruction in Spanish to intensive bilingual instruction which indicated that, de facto, the bilingual program employed in the school district sampled most clearly approximated Model I.

RESEARCH QUESTIONS

The major thrust of this investigation was directed toward answering three specific questions:

1. Do Mexican American students enrolled in a bilingual program at grade levels 3 and 5 attain significantly different scores on the Spanish versus English GTA reading test?
2. Do Mexican American males enrolled in a bilingual program at grade levels 3 and 5 attain significantly different scores than females on the Spanish versus English GTA reading test?
3. Do Mexican American students enrolled in one kind of bilingual program in grade levels 3 and 5 attain significantly different scores than Mexican American students in another kind of bilingual program (comparison group) on the Spanish versus English GTA reading test?

In answering these questions, subtest differences (i.e., vocabulary, level of comprehension, speed of comprehension) as well as subtest contributions to the total GTA battery score were also investigated and reported.

A secondary purpose of the study, based on the availability of data, was to determine the concurrent validity between the English GTA and the Comprehensive Tests of Basic Skills (CTBS), form Q, reading subtest. This comparison should be particularly useful since the CTBS has been used extensively to determine the general achievement and reading performance of Mexican American students.

METHODOLOGY

Because of practical limitations in data collection, the research employed a randomized control group post test only design. The sample was drawn from students in grades 3, 5, and 8 in order to determine whether the fit, or lack thereof, between English and Spanish GTA scores was attenuated across grade levels. In this respect, it would have been especially helpful to have had scores for grades K or 1, but this was precluded by the constraints of other priorities in conducting a large scale longitudinal evaluation of which this study is a small part. (3)

Research Design and Sample

The site of the study was a large, predominantly Mexican American, urban school district in South Texas. The sample for grades 3 and 5 was drawn from elementary schools within the district participating in two large federal programs designated as treatment and comparison groups. The treatment sample consisted of students participating in a National Institute of Education (NIE) program intended to reduce linguistic and cultural incompatibilities and thereby to promote significant academic growth. The comparison group consisted of students involved in a Title VII bilingual-bicultural program. The exact nature of the differences between the treatment and comparison bilingual programs cannot be further specified. Both programs, however, seem to have been based de facto on Model I as described in the introductory section of this paper.

Students in the grade 3 and 5 samples were randomly selected from classrooms designated bilingual and from a pool of students obtained on the basis of teacher and principal recommendations concerning pupil fluency in Spanish. In all likelihood, the key criterion employed in designating students for inclusion in these programs was oral Spanish fluency. The total grade 3 sample obtained was N=84, while the total grade 5 sample was N=76, with equal cell frequencies in treatment and comparison groups by sex. These two samples served as the major focus for the data analyses performed in this study.

In addition, a grade 8 sample of N=96 was also randomly selected from three classrooms each in treatment and comparison middle schools. In this case, however, the comparison middle school did not have a bilingual program, utilizing instead traditional curricula and instructional methods, while the treatment school was also part of the NIE program and utilized bilingual-bicultural curricula, self-paced individualized and small group instructional methods. For this reason, the 8th grade sample was utilized only in the secondary portion of this project, which tested the concurrent validity between the English GTA and the CTBS, form Q, reading subtest. All

schools sampled were similar in terms of ethnicity (Mexican American), socioeconomic status (lower SES), and achievement. (3)

INSTRUMENTATION

Guidance Testing Associates Reading Tests

The Guidance Testing Associates reading tests were selected for this study because they offered parallel versions of the same instrument specifically tailored to Mexican American youth in South Texas. Validation of the tests was accomplished by the GTA group centered at the University of Texas, Austin, and included students from forty-three states, Puerto Rico, and various Spanish speaking countries. (9) The majority of students in the norming group, however, was drawn from Texas public schools.

Parallel versions of the English and Spanish GTA tests consist of a pre and post test of reading, each of which includes five levels of difficulty: Level I - ages 6 and 7, second semester of grade 1 and beginning of grade 2; Level II - ages 7-8, second semester of grade 2, and grade 3; Level III - ages 9-11, grades 4-6; Level IV - ages 12-14, grades 7-9; Level V - ages 15-18, grades 10-12 and the beginning of 13. The pre test version (form CE) was used in this study (the pre test form CE differs only in that the order of items varies from the post test form DE).

Level II was administered to the third grade sample, Level III to the fifth grade, and Level IV to the eighth grade sample. Each level contains three subtests: 1. Vocabulary; 2. Level of Comprehension; and 3. Speed of Comprehension.

To establish reliability of the GTA tests, intercorrelations were run among the three subtests and the cumulative total on both the English and Spanish versions for the three grades under consideration: 3, 5, and 8. Coefficients of correlation for Vocabulary with Speed, Vocabulary with Level, and Speed with Level for grade three were .55, .58, and .80; grade five correlations were .76, .76, and .79; and grade eight correlations were .79, .71, and .76. Spanish intercorrelations for the same subtest were .34, .60, and .62 for grade three; and .63, .68, and .74 for grade five; and for grade eight they were .64, .73, and .61 respectively.

Correlation between the pre and post forms of the English GTA for grade three Vocabulary subtest was .66, Speed .87, Level .74, and Total .89; grade five correlations were .89, .84, .81, and .95; and grade eight correlations were .76, .77, .80, and .91 respectively. Coefficients of correlations on the Spanish forms were Vocabulary .66, Speed .39, Level .65,

and Total .71 for grade three; the grade five correlations were .85, .77, .80, and .90. Correlations for adjacent levels were also reported for the English GTA. Grade three to grade four correlations were for Vocabulary .49, Speed .72, Level .73, and Total .77; grade five to six correlations were .71, .81, .80, and .89; and, grade eight to nine were .85, .73, .73, and .90.

Validity of the GTA reading test had been established concurrently by comparing it to various other tests. The GTA Level II for grade three was compared with the Otis Quick Scoring Mental Ability Test and correlations were .51 for verbal-numerical, .75 for nonverbal, and .73 for Total. Correlations with the Metropolitan Reading Test were .68 for Level of Comprehension to reading and .65 for Vocabulary to word knowledge. Stanford Achievement Primary II and GTA correlations were .80 for Vocabulary, .51 for Speed, .72 for Level, and .78 for Total.

The Level III GTA for grade five was compared to the School and College Ability Test (SCAT). The reported correlations were .75 for Verbal-Verbal, .45 for Nonverbal-Total, .74 for Numerical-Quantitative, and .76 for Total-Total. The GTA Level IV, applicable to grade eight, correlations with the SCAT were reported to be .80, .41, .76, and .81 respectively. Correlations for Level IV with the California Mental Maturity and Achievement Test (N=328) were reported to be .75, .48, .56, and .71; and correlation with the California Achievement Test were GTA Vocabulary - CAT Vocabulary .64 and GTA Level to CAT Comprehension .68. Finally, correlations of Level III to the STEP reading were .70 for Vocabulary, .77 for Speed, .82 for Level, and .83 for Total. It appears, then, that the GTA generally measureswhat the other tests are measuring, but it has some (approximately one-third) unique variance also.

Comprehensive Tests of Basic Skills

The CTBS, form Q, was utilized in the secondary portion of this investigation for the reasons noted previously. This instrument, used extensively with Mexican American children, was standardized on a large national sample of students from grades 2 through 10, randomly selected from all regions and states of the United States. The sample included public and private students proportionate in number to actual enrollments. The publishing company, McGraw-Hill, reported a high degree of instrument reliability for both subtest as well as total scores. In terms of the Kuder-Richardson formula 20 for measuring internal consistency, the reliability of the CTBS ranged between .75 and .96 over all grade levels with most of the coefficients being over .80. (5)

Procedure

The tests were administered during April and May 1975. The order of the test administration was the English GTA followed one week later by the CTBS and three weeks later by the Spanish GTA. The period of time between the English and Spanish GTA was five weeks. Since there was no variation in the order of test administration among the sample, sequence bias is possible between the GTA tests. However, it must be noted that the five week period between GTA tests mitigated against correct item recall. Even assuming sequence bias, it would favor better performance on the Spanish GTA.

Analytical Techniques

The statistical tool selected to answer the three major research questions posed in this study was dummy variable regression analysis since it allowed the researchers to combine categorical variables (sex and group) with interval variables (English and Spanish GTA reading scores). (12) Pearson correlation was also used to derive a measure of association between the English and Spanish GTA and between the English GTA and the CTBS, form Q, reading subtest.

FINDINGS RELATED TO RESEARCH QUESTIONS

Four separate analyses are presented for each grade level: one for the total battery score comparison between the English and Spanish GTA and one each for the three subtest components (Vocabulary, Level of Comprehension, and Speed of Comprehension) which comprise the total battery. In each of the analyses the score obtained on the Spanish version of the GTA reading test or subtest served as the criterion variable, while group, sex, and the score obtained on the English version of the GTA reading test or subtest served as the independent variables. Findings for each of the analyses are presented graphically and by means of an ANOVA source table in which the mean square values were derived directly from the R^2 values in the dummy variable equation and each corresponding F value computed.

Grade 3

Since the grade 3 children had had formal bilingual instruction in first and second grade, it was expected that they would perform relatively well on the Spanish reading test. They were not, however, receiving concentrated training in Spanish reading during the year in which they were tested. Rather, they were receiving limited instruction in Spanish.

Total Battery Scores

As revealed in Table 1, the Mexican American third grade pupils achieved significantly different mean total battery scores on the Spanish and English versions of the GTA reading test. Contrary to what might have been expected from the review of the literature, the reported means indicated that students in the sample scored significantly higher on the English total battery (ETOT) as opposed to the Spanish version (STOT). There was also a marked difference by sex in performance on the Spanish version of the test (females scored 11.63 points higher than males). Within the comparison group, however, females scored higher than males in both the English and Spanish versions of the test, but the difference in performance for males and females was consistent across the two language versions of the test.

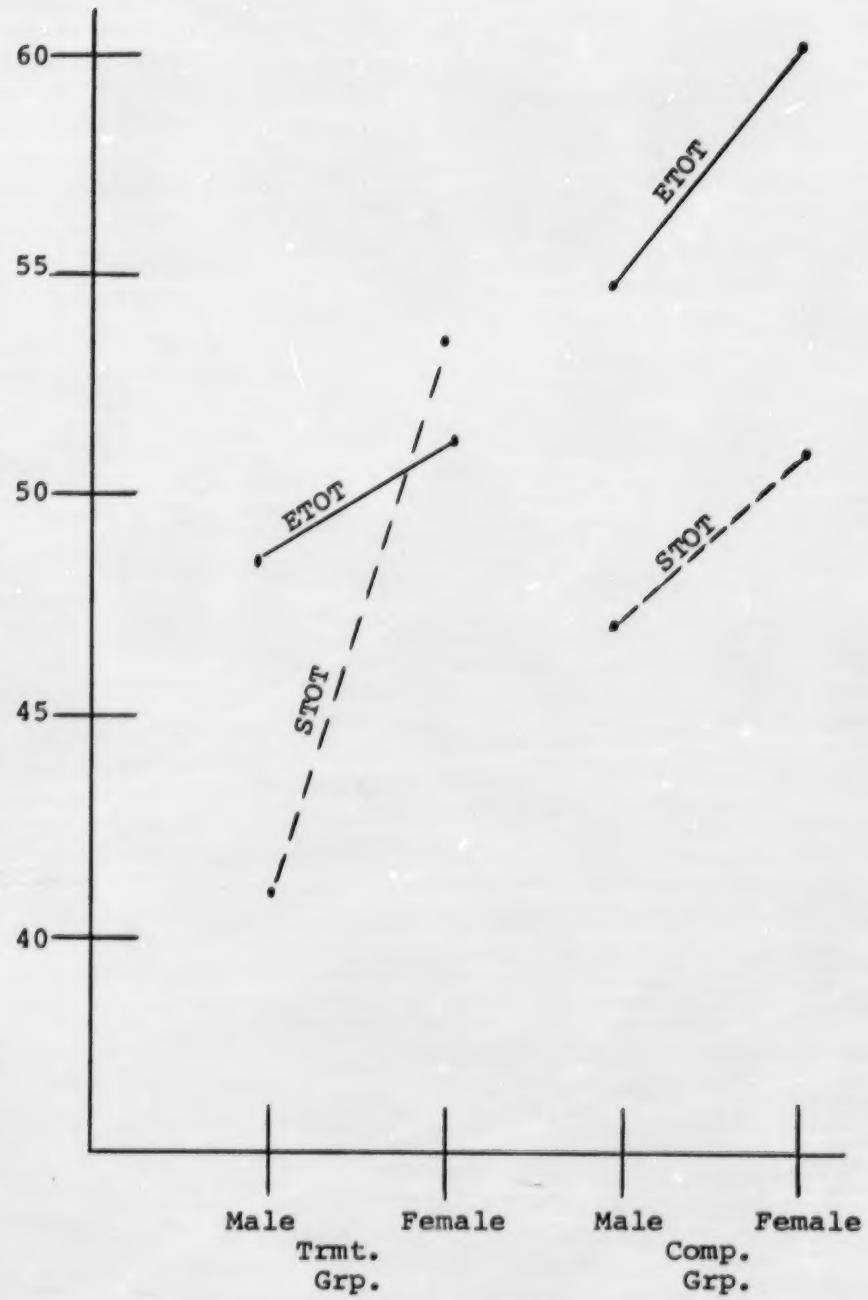
Table 1: ANOVA Source Table for Spanish Total Battery by Group and Sex: Grade 3

SOURCE	MS	DF	F
Saturated Model	.1044	7	14.108***
Additive Model	.2350	3	30.519***
Sex	.0148	1	1.922
Grp. controlling for sex	.00864	1	1.123
ETOT controlling for sex, grp.	.68156	1	88.583***
Interaction	.00642	4	.834
Grp. X sex	.00013	1	.017
ETOT X sex	.00857	1	1.088
ETOT X grp.	.00013	1	.017
ETOT X grp. X sex	.01705	1	2.216
Residual	.007694	76	

	TRMT. GRP.		COMP. GRP.	
Sex	ETOT	STOT	ETOT	STOT
M	48.72	41.90	54.91	47.24
F	51.05	53.53	59.96	50.86

***p<.001

Figure 1 : Total Battery Means by Group and Sex: Grade 3



Vocabulary Scores

As indicated in Table 2, students attained significantly different scores on the Spanish and English versions of the GTA vocabulary subtest. As evidenced by the means, students scored significantly higher on the English as opposed to the Spanish version of the subtest. There was a significant three-way interaction effect for group X sex X English Vocabulary score (EVOC), but no main effects for group or sex.

Inspection of Figure 2 reveals the nature of this interaction. While females scored higher than males on both the English and Spanish versions of the GTA vocabulary subtest in both the treatment and comparison groups, the size of the male-female differences was not consistent across treatment and comparison groups. While the difference in performance on the EVOC and Spanish Vocabulary (SVOC) subtests was approximately the same for males and females in the comparison groups, the difference in performance between the EVOC and SVOC scores was considerably smaller for females than males in the treatment group. Females in the treatment group scored only slightly higher than males on the EVOC subtest but significantly higher than males in the SVOC subtest.

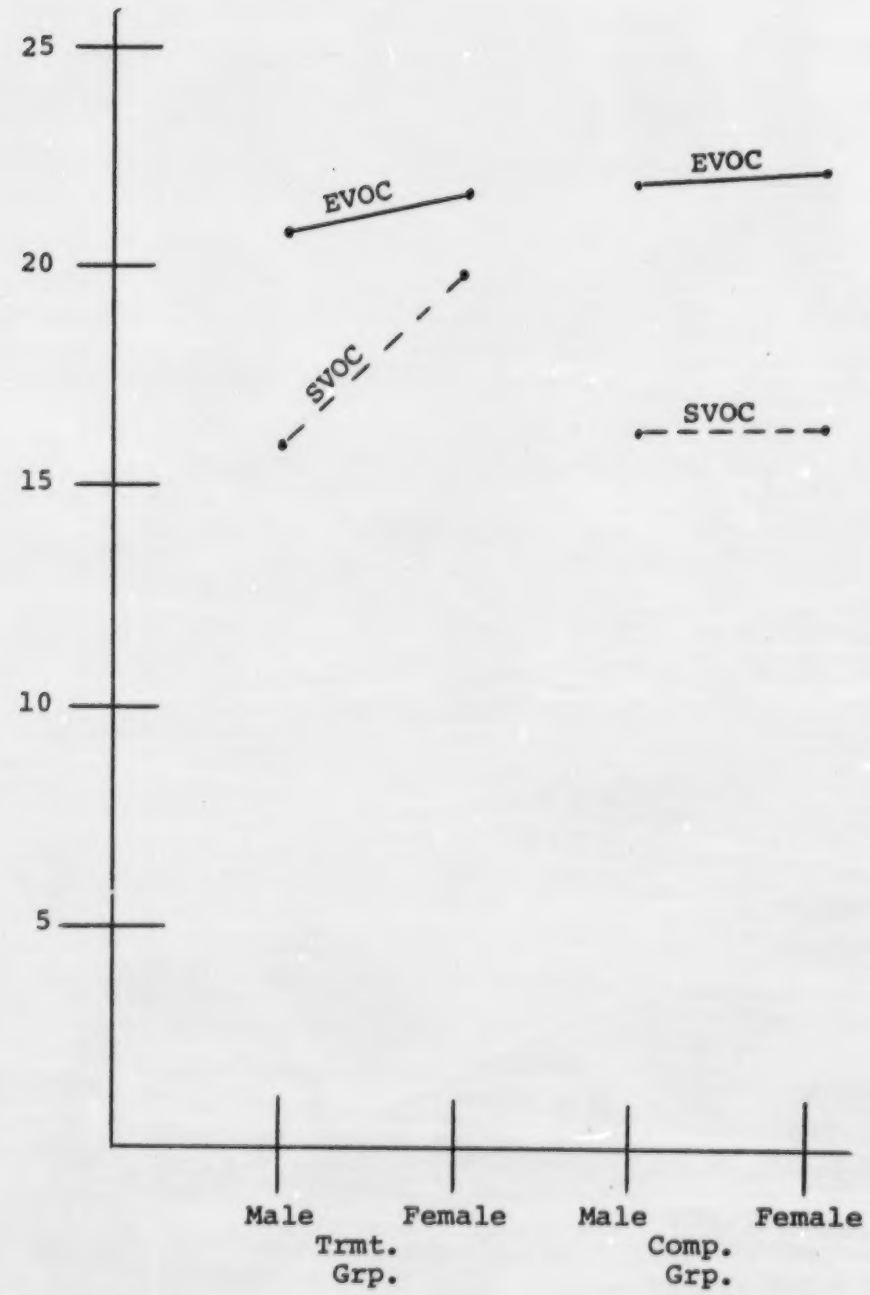
Table 2: ANOVA Source Table for Spanish Vocabulary and Means by Group and Sex: Grade 3

SOURCE	MS	DF	F
Saturated Model	.0808	7	14.175***
Additive Model	.1745	3	30.614***
Sex	.01377	1	2.415
Grp. controlling for sex	.01003	1	1.759
EVOC, controlling for sex, grp.	.49963	1	87.654***
Interaction Model	.01053	4	1.847
EVOC X sex	.01000	1	1.754
EVOC X grp.	.00113	1	0.198
Sex X grp.	.00002	1	0.004
EVOC X sex X grp.	.03097	1	5.433***
Residual	.00571	76	

	TRMT. GRP.		COMP. GRP.	
Sex	EVOC	SVOC	EVOC	SVOC
M	20.86	16.14	21.95	16.33
F	21.76	19.95	22.24	16.43

***p<.001

Figure 2 : Vocabulary Subtest Means by Group and Sex: Grade 3



Level of Comprehension

In the second subtest analysis it was found that all three independent variables of sex, group and level of English Comprehension (ELCOMP) had a main effect on the criterion variable, Level of Spanish Comprehension (SLCOMP); however, no interaction effects were detected (see Table 3). These results are illustrated graphically in Figure 3. As in previous analyses, the mean scores of females on the GTA Level of Comprehension was higher than the mean scores of males on both English and Spanish in the comparison as well as treatment groups. Furthermore, students in both the treatment and comparison groups scored significantly higher on the English as opposed to the Spanish versions of subtest II. The comparison group attained higher mean scores than the treatment group on both the English and Spanish versions of the GTA Level of Comprehension subtest.

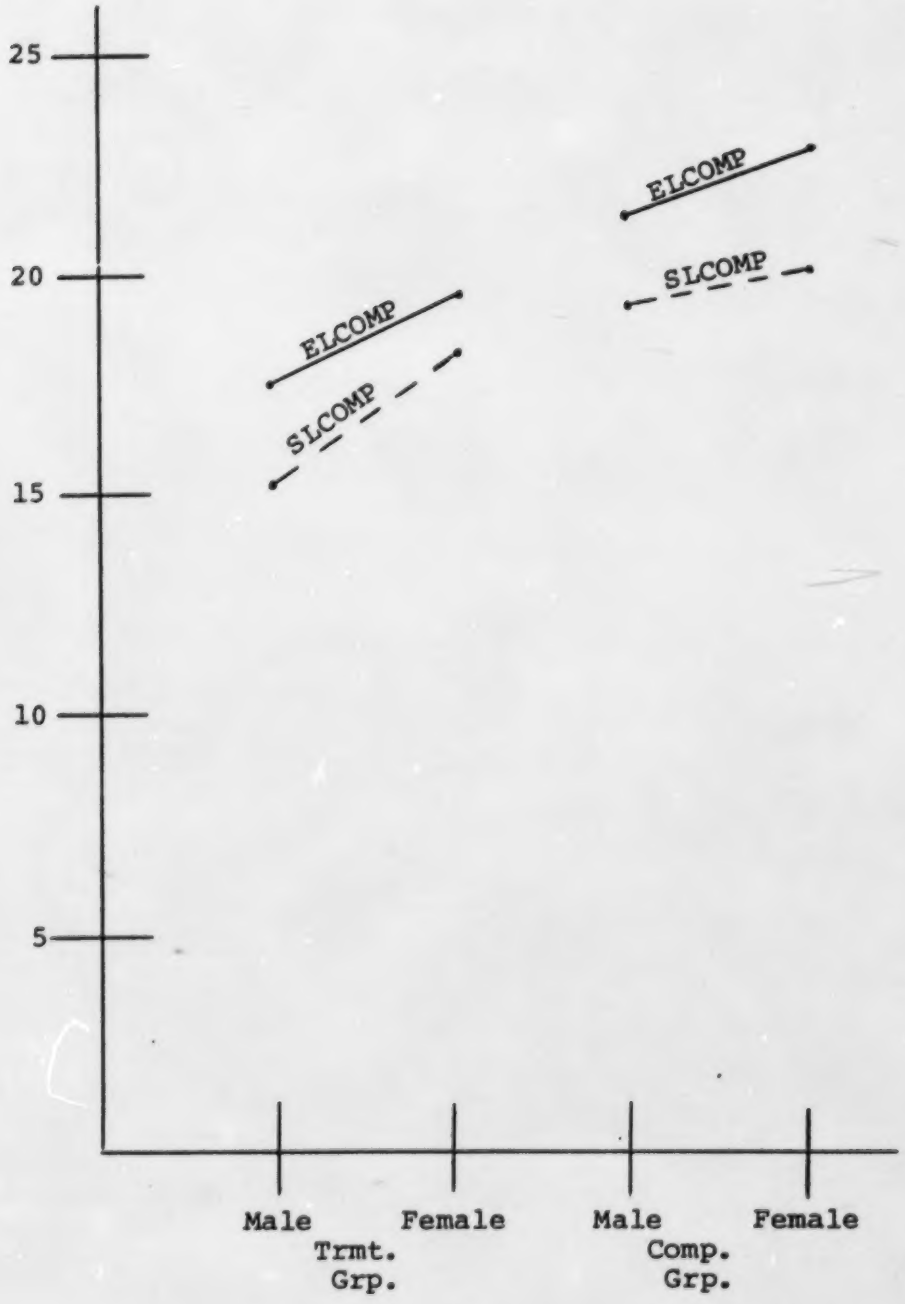
Table 3: ANOVA Source Table for Spanish Level of Comprehension by Group, Sex: Grade 3

SOURCE	MS	DF	F
Saturated Model	.0961	7	22.348***
Additive Model	.2195	3	51.046***
Sex	.01406	1	3.269*
Grp. controlling for sex	.03677	1	8.551***
ELCOMP controlling for sex, grp.	.06765	1	141.314***
Interaction	.0036	4	0.837
ELCOMP X sex	.0002	1	0.046
ELCOMP X grp.	.00123	1	0.286
Sex X grp.	.00325	1	0.756
ELCOMP X sex X grp.	.00958	1	2.228
Residual	.0043	76	

	TRMT. GRP.		COMP. GRP.	
Sex	ELCOMP.	SLCOMP.	ELCOMP.	SLCOMP.
M	17.69	15.19	21.52	19.48
F	19.62	18.29	22.91	20.24

*p<.05
***p<.001

Figure 3 : Level of Comprehension Subtest Means by Group and Sex : Grade 3



Speed of Comprehension

In the final subtest analysis, two of the independent variables, sex and Speed of English Comprehension, were found to have a significant main effect on the criterion variable, Speed of Spanish Comprehension (SSCOMP), as illustrated in Table 4. No significant interaction effect was detected. Figure 4 graphically illustrates some interesting differences involving the three independent variables. The nature of the significant main effects and the direction of the non-significant interaction effects were somewhat different from that found in the previous grade 3 analyses.

Table 4: ANOVA Source Table for Spanish Speed of Comprehension by Group, Sex: Grade 3

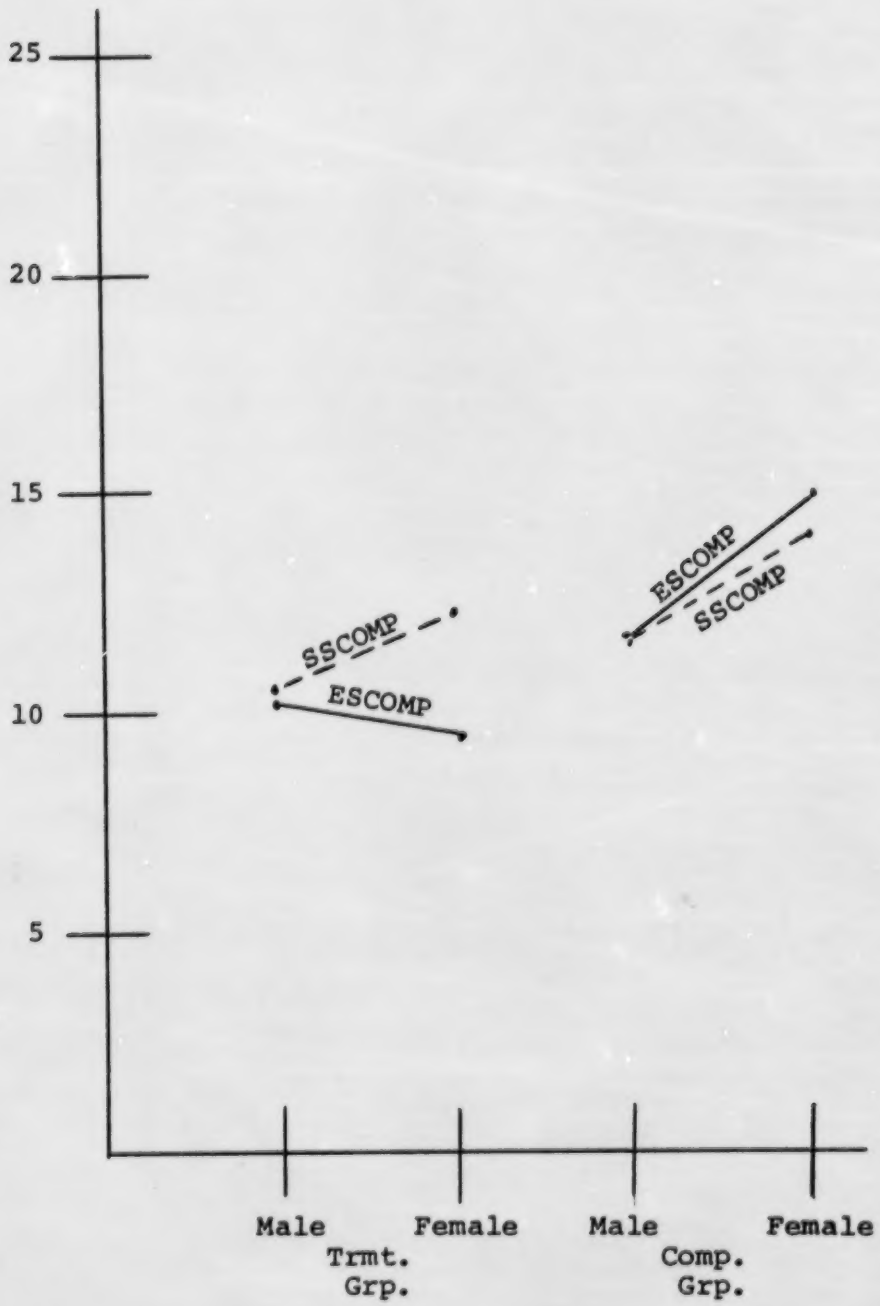
SOURCE	MS	df	F
Saturated Model	.0909	7	18.938***
Additive Model	.2055	3	42.813***
Sex	.03685	1	7.677***
Grp. controlling for sex	.01403	1	2.923
ESCOMP controlling for sex, grp.	.56573	1	117.869***
Interaction	.0050	4	1.04
ESCOMP X sex	.00139	1	1.35
ESCOMP X grp.	.00420	1	0.29
Sex X grp.	.00650	1	0.88
ESCOMP X sex X grp.	.00769	1	1.60
Residual	.0048	76	

	TRMT. GRP.		COMP. GRP.	
Sex	ESCOMP.	SSCOMP.	ESCOMP.	SSCOMP.
M	10.19	10.57	11.67	11.43
F	9.67	12.29	14.81	14.19

***p<.001

The most notable difference was that these Mexican American students did not consistently attain higher mean scores on the English version of the subtest as compared to the Spanish version. That is, students within the treatment group attained higher mean scores on the Spanish GTA Speed of Comprehension test, while students in the comparison group still attained higher scores on the English version of this subtest.

Figure 4: Speed of Comprehension Subtest by Group and Sex: Grade 3



In addition, while females attained higher mean scores than males on the Spanish version of the GTA Speed of Comprehension subtest in both the treatment and comparison groups, they attained a lower mean score than males on the English version of the subtest within the treatment group though continuing to attain higher mean scores on this language version than the males within the comparison group.

Summary and Discussion

In terms of the research questions posed, the data revealed that Mexican American students in a third grade Model I bilingual program, who were selected for inclusion in the program primarily on the basis of oral Spanish fluency, attained significantly higher mean scores on all English subtests of the GTA reading test than on the Spanish version of the GTA, with the one exception: treatment students attained a higher mean score on the Spanish Speed of Comprehension subtest than on the English version of that subtest. For the total GTA battery, mean scores for all students were significantly higher on the English as compared to the Spanish edition of the test. These significant differences cast considerable doubt on the commonly held assumption that Spanish language tests are ipso facto superior to English language tests for Mexican American, Spanish-surnamed bilingual pupils. Clearly, both the level and type of Spanish fluency used as a criterion for selection into the program, as well as the intended learning outcomes (i.e., the model) of the instructional program, must be assessed before determining the appropriateness of testing academic achievement exclusively in Spanish or in interpreting the meaning of differences between Spanish and English test results for students in a bilingual program.

The generally lower mean test scores attained by the students on the GTA may be reflective of one or several factors at work when Model I is used as a basis for bilingual instructional programs. First, since selection into the program is based on oral Spanish fluency and intensive instruction in Spanish terminates at or about grade 2 while instruction in English is accelerated, students simply lack the formal training and practice necessary to score as well on a Spanish version of an achievement test as they do on an English version of that same test. Secondly, when Model I is used as the basis of instruction, as Spanish is de-emphasized and training in English occurs at an ever accelerated pace, the teacher may serve as a model giving the students cues that English is the legitimate language of the schools and of the dominant society regardless of their own participation in a bilingual program. Consequently, the students perceive that the appropriate language in which to achieve in school is English and give ever lessening attention to the acquisition of the more formal Spanish proficiency measured by the GTA. Given either or both of the above factors, testing the reading achievement of these third

grade Mexican American students exclusively in Spanish would have significantly underestimated their level of achievement vis-a-vis other students and could have detrimentally affected their educational opportunities.

In addition to the important language differences noted, significant main effects for sex were found on two of the subtests, Level of Comprehension and Speed of Comprehension, with the direction of this sex difference being supported by trends in the Vocabulary subtest and Total Battery scores. This sex difference was reflected in consistently higher mean scores for females than males on the total battery and subtests of both language versions of the GTA with the one exception that treatment males scored slightly higher than females on the English Speed of Comprehension test.

One interesting interaction effect, related to the above sex differences, was also manifest in the data. That is, there was a significant three-way interaction effect for English Vocabulary X sex X group when regressed on the criterion variable of Spanish Vocabulary. An examination of this effect revealed that the difference in Spanish and English Vocabulary scores was constant across the categories male-female in the comparison group, while the difference in Spanish and English Vocabulary scores was considerably smaller for females than for males within the treatment group. That is, treatment females scored significantly higher in Spanish Vocabulary than treatment males, but only slightly higher in English Vocabulary. These findings with regard to the main and interaction effects of sex warrant further investigation to determine the exact nature of such differences by sex.

Finally, it can be noted that while only one significant difference was found between third grade Mexican American students in the comparison as contrasted with the treatment groups, Figures 2, 3, and 4 show a clear trend toward higher performance by comparison group students in both English and Spanish. The one exception to this was the vocabulary subtest in which treatment females scored significantly higher than those in the comparison group. This also accounts for the same phenomenon when looking at the total battery scores reported in Table 1 and illustrated in Figure 1. While it should be cautioned that the sample selection procedures do not guarantee that students in the treatment and comparison bilingual programs were equal in mean scores initially, the above findings concerning noticeable differences in performance between students in these two groups highlight the need to identify the particular components of each bilingual program and to determine their differential impact on Mexican American students.

Grade 5

The fifth grade children had received a full bilingual program only one year, when they were second graders. After that,

Spanish was used in classroom management and in some instructional programs. Some lessening of Spanish proficiency could thus be expected.

Total Battery Scores

As revealed in Table 5 and graphically illustrated in Figure 5, these Mexican American fifth grade students attained significantly higher mean scores on the English version of the GTA total reading battery than on the Spanish version. While no statistically significant effects were found for sex or group, Figure 6 indicates that females generally attained higher mean scores than males with the exception that treatment females scored slightly lower than treatment males on the English GTA total battery. In addition, students in the comparison group scored generally higher on both the English and Spanish versions of the GTA than students in the treatment group. No statistically significant interaction effects were present.

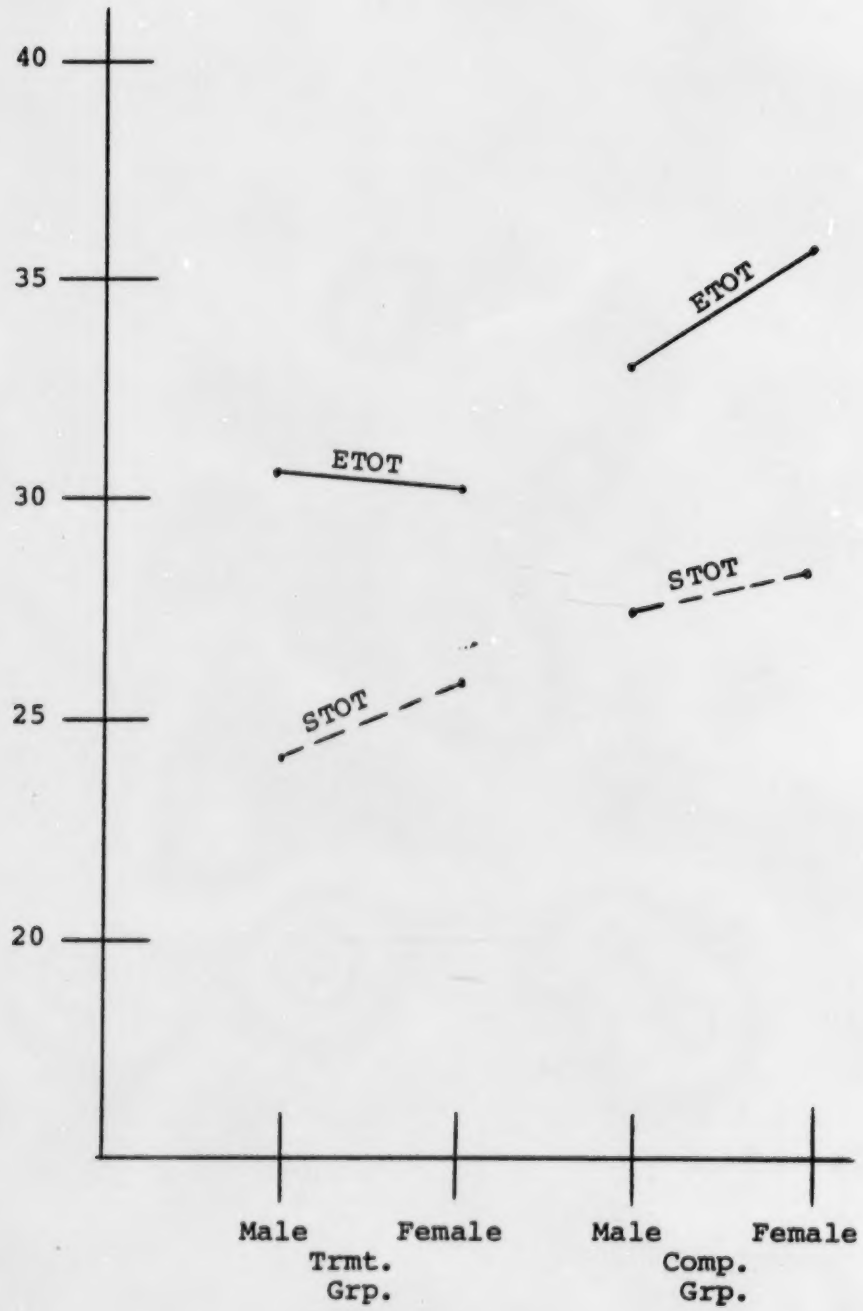
Table 5: ANOVA Source Table for English Total Battery by Group and Sex: Grade 5

SOURCE	MS	df	F
Saturated Model	.06880	7	9.025***
Additive Model	.15441	3	20.255***
Sex	.00322	1	0.422
Grp. controlling for sex	.01926	1	2.526
ETOT controlling for grp. and sex	.44074	1	57.815***
Interaction Model	.00459	4	0.603
Grp. X sex	.00594	1	0.779
ETOT X sex	.00113	1	0.148
ETOT X grp.	.01127	1	1.478
ETOT X grp. X sex	.00005	1	0.066
Residual	.00762	68	

	TRMT. GRP.		COMP. GRP.	
Sex	ETOT	STOT	ETOT	STOT
M	31.68	24.27	33.16	27.47
F	31.32	25.84	35.58	28.53

***p<.001

Figure 5: Total Battery Means by Group and Sex: Grade 5



Vocabulary Scores

The analysis of the vocabulary subtest, reported in Table 6, revealed main effects for all three independent variables as well as a two-way interaction effect for group X sex. Examination of the means reported in Table 6, as well as the graphic representation in Figure 6, indicate that Mexican American fifth grade students attained a significantly higher mean score on the English Vocabulary version of the GTA subtest than on the Spanish version. In addition, the mean score of females tended to be higher than that of males on both the Spanish and English version with the exception that treatment females attained slightly lower mean scores on the English Vocabulary subtest than males. Further, students in the comparison group achieved higher mean scores on both the English and Spanish versions of the subtest than students in the treatment group.

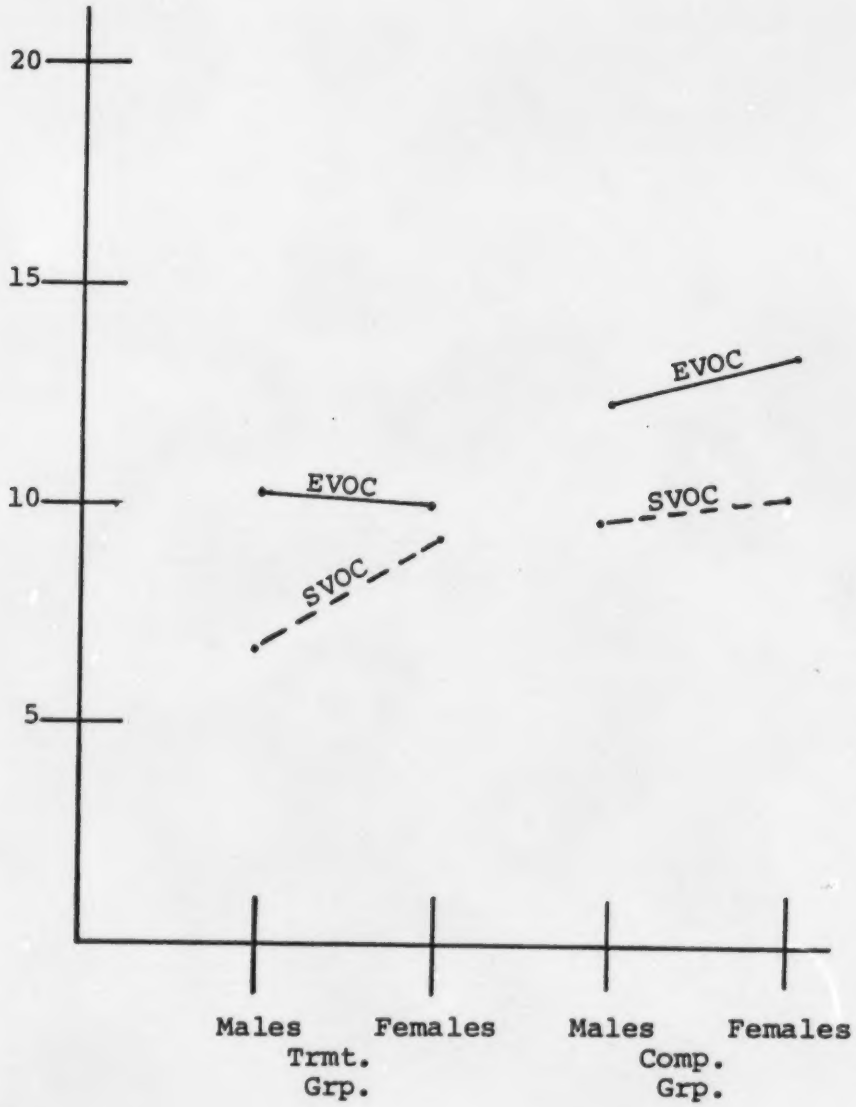
Table 6: ANOVA Source Table for Spanish Vocabulary by Group and Sex: Grade 5

SOURCE	MS	df	F
Saturated Model	.0449	7	9.707***
Additive Model	.17624	3	38.065***
Sex	.01475	1	3.186*
Grp. controlling for sex	.05193	1	11.216***
EVOC controlling for sex, grp.	.20655	1	44.611***
Interaction Model	.07007	4	15.134***
EVOC X sex	.00207	1	0.447
EVOC X grp.	.01081	1	2.335
Grp. X sex	.01544	1	3.335*
EVOC X grp. X sex	.00463	1	0.754
Residual	.00463	68	

	TRMT. GRP.		COMP. GRP.	
Sex	EVOC	SVOC	EVOC	SVOC
M	10.16	6.79	12.42	9.68
F	9.95	9.05	13.47	10.16

*p<.05
***p<.001

Figure 6: Vocabulary Subtest Means by Group and Sex: Grade 5



The interaction effect for group X sex was due to the differences in mean scores between the English and Spanish versions of the subtest. These were constant across sex in the comparison group, whereas these differences were not constant in the treatment group. That is, treatment females scored slightly lower than males on the English Vocabulary subtest, but significantly higher than males on the Spanish version of the subtest.

Level of Comprehension

Results for Subtest II are displayed in Table 7 and Figure 7. These indicate that there was no significant interaction effect present and that only English Level of Comprehension evidenced a main effect. As indicated by the means reported in Table 7, this main effect was due to these fifth grade Mexican American students achieving higher mean scores on the English Level of Comprehension subtest than on the parallel Spanish version. While not statistically significant, it should also be noted that a moderate but clear reversal of sex effect occurred: males generally scored higher than females on both versions of this subtest with the exception that comparison females attained a higher mean score on the Spanish version of the test than males.

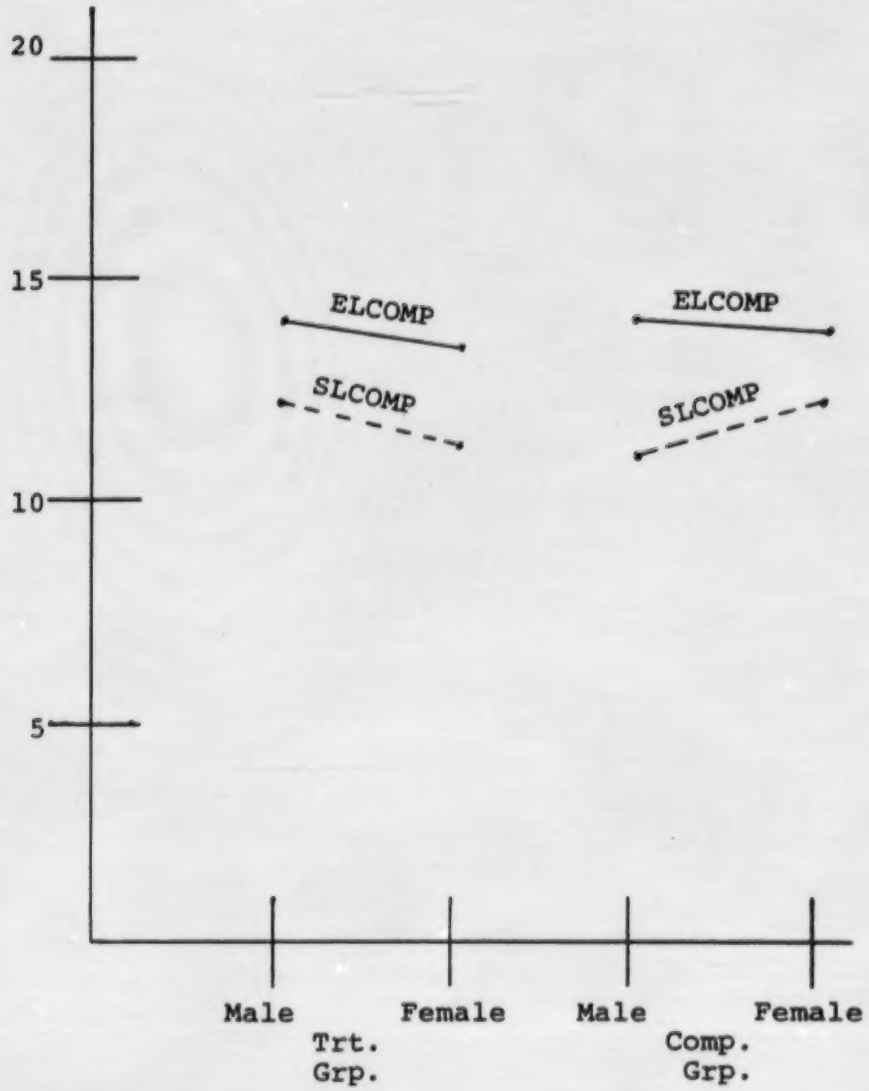
Table 7: ANOVA Source Table for Level of Comprehension Spanish by Group, Sex, and Comprehension English: Grade 5

SOURCE	MS	df	F
Saturated Model	.05314	7	5.754***
Additive Model	.11862	3	12.844***
Sex	.00012	1	.013
Grp. controlling for sex	.00024	1	.026
ELCOMP controlling for sex, grp.	.35550	1	38.493***
Interaction Model	.00403	4	.436
Grp. X sex	.00680	1	.736
ELCOMP X sex	.00563	1	.609
ELCOMP X grp.	.00033	1	.036
ELCOMP X grp. X sex	.00336	1	.364
Residual	.00924	68	

	TRMT. GRP.		COMP. GRP.	
Sex	ELCOMP	SLCOMP	ELCOMP	SLCOMP
M	14.05	12.32	14.16	11.11
F	13.63	11.42	13.95	12.26

***p<.001

Figure 7: Level of Comprehension Subtest Means by Group and Sex: Grade 5



Speed of Comprehension

Table 8 and Figure 8 illustrate findings of the analysis performed for Speed of Comprehension. These reveal a main effect for English Speed of Comprehension as well as a three-way interaction effect for English Speed of Comprehension X group X sex. As with all previous analyses, the main effect for language was that these Mexican American students attained significantly higher mean scores on the English version of the GTA Speed of Comprehension subtest than on the Spanish version. While neither effect was statistically significant, it can also be noted from Figure 8 that students in the treatment group tended to attain higher mean scores on both versions of this subtest than students in the comparison group. In addition, females tended to attain higher mean scores than males with the exception of lower performance by comparison females on Spanish Speed of Comprehension.

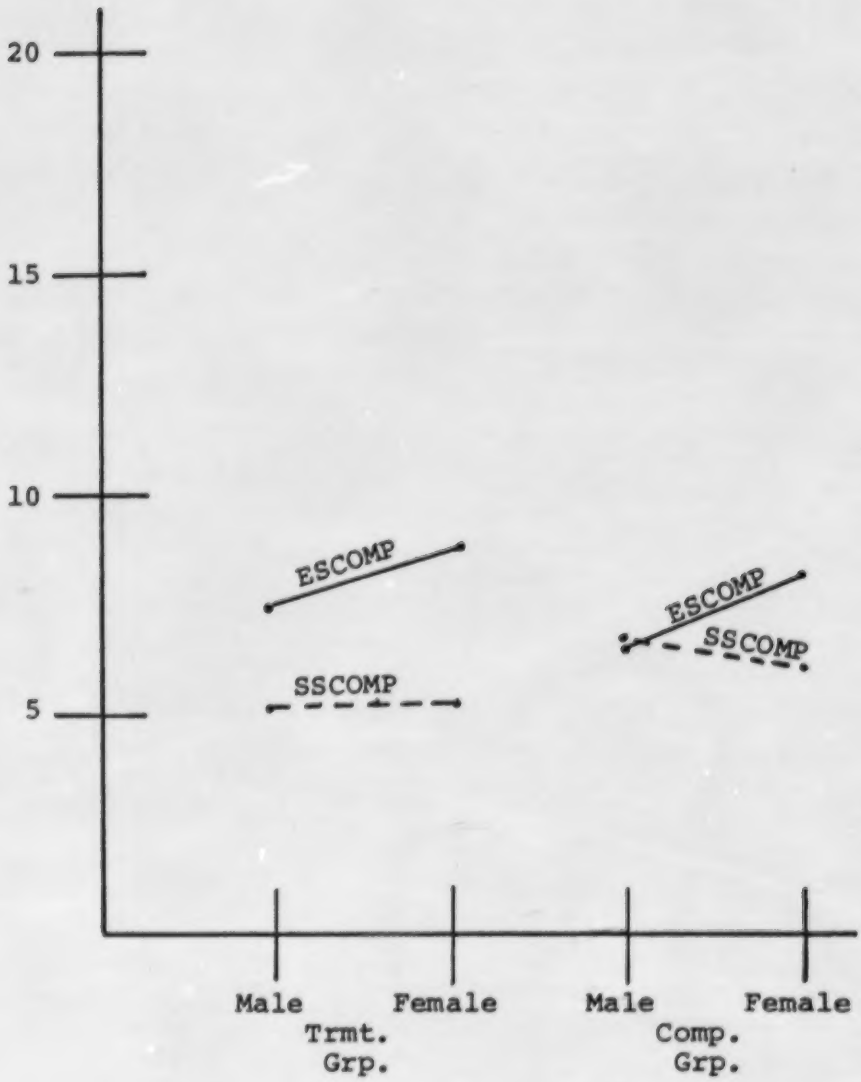
Table 8: ANOVA Source Table for Speed of Comprehension in Spanish by Group, Sex, and English Comprehension: Grade 5

SOURCE	MS	df	F
Saturated Model	.04533	7	4.515**
Additive Model	.08016	3	7.984**
Sex	.00086	1	0.087
Grp. controlling for sex	.03258	1	3.245
ESCOMP controlling for sex, grp.	.20702	1	1.913
Interaction Model	.01921	4	1.913
Grp. X sex	.00535	1	0.533
ESCOMP X sex	.00125	1	0.125
ESCOMP X grp.	.00508	1	0.506
ESCOMP X grp. X sex	.06514	1	6.488**
Residual	.01004	68	

	TRMT. GRP.		COMP. GRP.	
Sex	ESCOMP	SSCOMP	ESCOMP	SSCOMP
M	7.58	5.16	6.58	6.68
F	8.74	5.37	8.16	6.11

*p<.05
**p<.01

Figure 8 : Speed of Comprehension Subtest Means by Group and Sex: Grade 5



The three-way interaction effect, illustrated in Figure 8, was probably due to the difference in mean scores between the English and Spanish versions of the subtest. The difference between Spanish and English mean scores was fairly constant across sex in the treatment group. In the comparison group, however, comparison males achieved approximately the same mean scores on the English and Spanish versions of the subtest, while females attained a significantly lower mean score on the Spanish Speed of Comprehension subtest than on the English.

Summary and Discussion

Analyses performed on a sample of fifth grade Mexican American students in two bilingual programs, comparing performance on the English GTA reading test to performance on a parallel Spanish version of that test, with controls for sex and group, indicated that these students consistently achieved higher mean scores on the English as opposed to the Spanish version of the test and each of its component subtests. While statistical significance was evidenced only in the case of the Vocabulary subtest, there was also a marked trend for females to attain higher mean scores on both versions of the GTA reading test than males. This trend was reversed, however, in the Level of Comprehension subtest. Further, there was a trend, though not statistically significant, for comparison group students to achieve higher mean scores on both versions of the GTA with the one exception of the Speed of Comprehension subtest.

Only two interaction effects were detected: a group X sex interaction in the Vocabulary subtest and a group X sex X English Speed of Comprehension effect in the Speed of Comprehension subtest. Both of these were due to differential Spanish-English performance by males or females in either the treatment or comparison groups.

These sets of findings would again suggest that an a priori decision to test bilingual students in Spanish as opposed to English on the assumption that they would perform better in their native language is ill advised. The students' level of Spanish language proficiency, the nature of the bilingual program in which they are participating, the amount of Spanish instruction they have received, and the type of test to be administered must be carefully considered if a Spanish version of a test is not to have the reverse effect for which it is intended and penalize some bilingual Mexican American students.

Finally, it should be noted that differences in performance by sex and by treatment indicate the need for additional study to determine the precise manner in which the distinct components of different bilingual programs impact the school performance of bilingual Mexican American students. Elements within the programs may have greater appeal to one group or another, or may be more effective with some students than with others.

Comparison of Findings for Grades 3 and 5

The consistent finding for both grade 3 and 5 students was that their performance in an English version of the GTA was superior to their performance in the parallel Spanish version of that test. Similar sex and treatment effects were also found for both groups, indicating that the precise nature of these variables should be assessed and given further study.

Correlations of Reading Tests

This segment of the investigation focused on assessing 1. the relationship of the English and Spanish editions of the GTA at the third and fifth grade levels; and 2. the relationship of the GTA vocabulary and Level of Comprehension subtests to the CTBS vocabulary and comprehension subtests at the third, fifth, and eighth grade levels. It was posited that since the English and Spanish GTA are parallel editions, the correlations between the two would be high across the grade levels tested. Similarly, since the GTA Vocabulary and Level of Comprehension and CTBS Vocabulary and Comprehension subtests purported to measure the same reading domain, the correlations would be stable across grade levels tested, suggesting that the two should be interchangeable.

English-Spanish GTA Correlations

Correlation coefficients by GTA subtests for the third (N=84) and fifth grade sample (N=76) were derived and are presented in Table 9.

Table 9: English-Spanish GTA Correlations for Grades 3, 5

Subtests	3*	5*
EVOC-SVOC	.705	.503
ELCOMP-SLCOMP	.810	.596
ESCOMP-SSCOMP	.771	.429
ETOT-STOT	.836	.675

*all correlations $p<.001$

The third grade correlations indicated a relatively high degree of association between the English and Spanish GTA, which suggested that these tests are appropriate for this grade level. The fifth grade correlations, however, were moderate and were well below those obtained for the third grade sample, which may be a direct reflection of increased English language usage, lessened intensity of Spanish instruction, and/or Spanish language displacement. Indeed, it may be that the

quality and/or character of bilingual instruction students received was increasingly varied as students progressed from the early elementary to later elementary grades. In any case, the moderate correlations between the English and Spanish GTA tests at the fifth grade level raised serious questions regarding the propriety of bilingual testing in the later elementary grades.

English GTA-CTBS Correlations

To determine the degree of association between the English GTA and CTBS on comparable Vocabulary and Comprehension subtests, correlation coefficients were derived. The standardized means, standard deviations, and correlation coefficients for grades three, five, and eight are presented in Table 10.

Table 10: English GTA-CTBS Correlations: Grades 3, 5, and 8

Subtests	3*			5*			8*		
	$\bar{x}$	SD	r	$\bar{x}$	SD	r	$\bar{x}$	SD	r
GTA Vocab.	49.79	8.99		51.16	9.45		49.38	10.19	
to			.551			.605			.544
CTBS Vocab.	49.23	9.72		50.49	9.27		50.15	9.33	
GTA Comp.	49.70	9.53		49.36	12.85		50.32	10.56	
to			.475			.523			.523
CTBS Comp.	50.15	10.44		51.04	9.72		50.55	8.97	

*all correlations $p<.001$

By inspection, these data reveal that the concurrent validity between comparable GTA and CTBS subtests was moderate at the grade levels tested. The standardized means between the Vocabulary and Comprehension subtests for grades 3 and 8 varied less than one point, and for grade 5, less than two points. In effect, the sample of Mexican American bilingual students tested at grade 3, 5, and 8 obtained similar standardized scores on the GTA and CTBS at their respective grade levels. A t-test analysis revealed no significant differences between the GTA and CTBS subtests among the samples in the grade levels tested. The probabilities that these differences were due to chance on the GTA-CTBS Vocabulary subtest for grades 3, 5, and 8 were .58, .51, and .42 respectively, and for the Comprehension subtests .69, .20 and .82 respectively.

Although the analysis of GTA and CTBS selected subtests was exploratory, the wide usage of these tests among Mexican American students suggest that careful analysis of the psychometric properties is warranted in view of the moderate correlations obtained. The selection of either of these tests to determine

student achievement growth should also be weighted against program content students are receiving.

DISCUSSION AND IMPLICATIONS

This study was conducted to ascertain whether bilingual Mexican American students scored significantly different on parallel Spanish-English reading tests and to provide Spanish language test data. Students were selected from two different bilingual programs and on the basis of Spanish language fluency from grade 3 (N=84) and grade 5 (N=76). Additionally, contrasts between these students' scores and those of a grade 8 sample (N=96) on the English GTA and comparable CTBS subtest were examined. Correlation coefficients between the English GTA and CTBS vocabulary and comprehension subtests for grades 3, 5, and 8 were obtained.

Dummy variable regression analysis was used to examine categorical and interval data relationships. The data revealed that the grade 3 and 5 samples obtained significantly higher scores on all English GTA subtests as compared to the Spanish GTA. There were, in addition, several sex and group differences.

The correlations between the English and Spanish GTA subtests were relatively high at the grade 3 level and ranged from .71 to .84. The grade 5 correlations were moderate and ranged from .43 to .68. The correlations between the English GTA and CTBS on comparable subtests for grades 3, 5, and 8 were also moderate. There were no significant differences between these students' standardized English GTA or CTBS subtest scores.

This study raises serious questions about the commonly held assumption that bilingual students' performance on a Spanish test will be superior to that in English. The grades 3 and 5 English and Spanish test data demonstrated that testing exclusively in Spanish would have seriously and detrimentally underestimated the competencies and abilities of these bilingual-bicultural Mexican American students.

Many educators have seriously questioned and challenged the propriety of testing bilingual Mexican Americans in English and have called for "culturally relevant" and Spanish language testing. In many instances, such criticisms are valid and do not excuse the past or present injustices committed by inappropriate testing and deleterious use of test data. However, as suggested by this study, testing in Spanish may have as many dangers as testing in English. Clearly, inappropriate testing in whatever language can only result in limiting the educational and social opportunities of students. It may well be that multiple measures are essential to more accurately derive needed achievement data. (2)

In addition, the findings also suggested that educational practitioners should closely examine the quality of bilingual programs. If such programs are intended to foster Spanish and English language competencies, as those purported to do in this study, the data suggested they are falling short of their objective. Spanish language fluency of bilingual-bicultural students in a predominately Mexican American school district (90%), from which students in this study were selected, appeared to decrease considerably after early elementary grades. While administrators and designers of the bilingual programs from which students sampled in this study were selected may have labeled their programs "bilingual," the de facto use of a Model I program design in which Spanish language training received little intensive emphasis beyond grade 2, resulted in a dramatic displacement of Spanish language among Mexican American students within the school setting. A re-examination of these bilingual programs both with respect to their philosophical premises and their pedagogical practices appears warranted.

Lastly, this study suggested that there is a dire need for additional empirical studies in the area of Spanish-English language dynamics with respect to student achievement outcomes. Assessment of language dominance, pre and post English-Spanish testing,and controls of home environmental and school variables would provide much needed information toward understanding the numerous complexities of the educational process as it relates to Mexican American students.

BIBLIOGRAPHY

1. BUDD, WILLIAM C. Educators and cultural-fair intelligence tests. Journal of Educational Sociology :333-335. March 1954.

2. CERVANTES, ROBERT A. Problems and alternatives in testing Mexican American students. Integrated Education XII(3):33-37. May-June 1974.

3. _____ ; JONES, EARL ET AL. Second interim report of the Edgewood School Plan. San Antonio, Texas. January 15, 1975.

4. CHANDLER, JOHN T. & PLAKOS, JOHN. Spanish speaking pupils classified as educable mentally retarded. Sacramento, California, State Department of Education, 1967.

5. CTB/McGRAW-HILL. Comprehensive tests of basic skills, form Q, all levels: technical bulleting. Monterey, California, 1968.

6. GARCIA, JOHN. IQ = the conspiracy. Psychology Today 6(4):40-43, 92-94. September 1972.

7. GUERRA, MANUEL H. "Why Juanito doesn't read." CTA Journal. October 1975.

8. HERNANDEZ-CHAVEZ, EDUARDO ET AL. El lenguaje de los Chicanos: regional and social characteristics of language used by Mexican Americans. Rosslyn, Virginia, Center for Applied Linguistics, 1975.

9. MANUEL, HERSCHEL. Technical report: tests of general ability and tests of reading. Guidance Testing Associates, Austin, Texas, August 1967.

10. MERCER, JANE R. IQ = the lethal table. Psychology Today 6(4):44-47, 95-97. September 1972.

11. MORENO, STEVE. Problems related to present testing instruments. El Grito III(3):25-29. Spring 1970.

12. NIE, NORMAN ET AL. Statistical package for the social sciences, sixth edition. New York, McGraw-Hill, 1975.

13. PALOMARES, UVALDO H. Performance of Mexican American children on selected nonverbal tests of intelligence. in JOHNSON, HENRY S. & HERNANDEZ, M. (eds.) Educating the Mexican American. Valley Forge, Pennsylvania, Judson Press, 1970. pp. 155-159.

14. _____. A critical analysis of research on the intellectual evaluation of Mexican American children. May 1965. (ERIC ED 027097)

15. ROSEN, CARL L. & ORTEGO, PHILLIP D. Language and reading problems of Spanish speaking children in the Southwest. Journal of Reading Behavior. Winter 1969.

16. SANCHEZ, GEORGE I. Group differences and Spanish speaking children--a critical review. Journal of Applied Psychology (16):549-558. 1932.

17. WEBER, GEORGE. Uses and abuses of standardized testing in the schools. Washington, DC, Council for Basic Education, May 1974.